Biophysical Journal, Volume 99

Supporting Material

Predicting ion binding properties for RNA tertiary structures

Zhi-Jie Tan and Shi-Jie Chen

Supporting Material

Predicting ion binding properties for RNA tertiary structures

Zhi-Jie Tan[1] **and Shi-Jie Chen**[2*]
[1]Department of Physics and Key Laboratory of Artificial Micro & Nano-structures of Ministry of Education, School of Physics and Technology, Wuhan University, Wuhan, P.R. China 430072
[2]Department of Physics and Astronomy and Department of Biochemistry, University of Missouri, Columbia, MO 65211

[*]Email: chenshi@missouri.edu.

Tightly bound ion (TBI) model at all-atom level

The original Tightly bound ion (TBI) model was based on coarse-grained structures [1]-[4]. Here we develop a new TBI model that can treat all-atom structures.

For multivalent ions such as Mg^{2+} ions, the higher charge and thus the strong Coulomb interaction between the ions can cause strong coupling (correlation) between the ions. To treat such effects, according to the correlation strength, we classify the counterions into two types [1]-[6]: the strongly and weakly correlated ions, denoted as the tightly bound and diffusive ions, respectively. Correspondingly, the space occupied by these ions are denoted as the tightly bound region and diffusive region, respectively; see Fig. 1 for examples of the tightly bound regions for the different structures. For the diffusive ions, we use the mean-field (PB) approach, while for the tightly bound ions, we must consider discrete modes of ion distributions to account for the ion-ion correlation effect. For a mixed monovalent and multivalent ion solution, because the correlation effect for the monovalent ions is negligible, monovalent ions can be treated as diffusive ions that form a background for multivalent ion binding.

Evaluation of ion correlation requires enumeration of the discrete distributions of the tightly bound ions and calculation of the free energy of the system for each given ion distribution mode. To enumerate the discrete ion distributions, for an N-nt RNA, we divide the tightly bound region into N cells, each around a phosphate. There exist a large number of "modes" ("binding modes") of ion distributions represented as the different ways to partition the tightly bound ions into the different cells. The total partition function Z is given by the summation over all the possible binding modes M:

$$Z = \sum_M Z_M. \tag{1}$$

Z_M is the partition function for a given binding mode M [1]-[4]

$$Z_M = Z^{(id)} \ (c_z)^{N_b} \left(\int \prod_{i=1}^{N_b} d\mathbf{R}_i \right) \ e^{-(\Delta G_b + \Delta G_d + \Delta G_b^{pol})/k_B T}, \tag{2}$$

where $Z^{(id)}$ is the partition function for the uniform ion solution (without the polyelectrolyte). c_z is the concentration of z-valent counterions. N_b is the number of tightly bound ions and $\int \prod_{i=1}^{N_b} d\mathbf{R}_i$ is the volume integral for the N_b tightly bound ions.

ΔG_b in Eq. 2 is the mean Coulombic interaction energy between all the charge pairs (including the phosphates and the tightly bound ions) in the tightly bound region. ΔG_b is calculated as the summation of pair-wise potentials of mean force [1]-[4]

$$\Delta G_b \simeq \sum_i \Phi_1(i) + \sum_{ij} \Phi_2(i, j), \tag{3}$$

where $\Phi_1(i)$ is the potential of mean force between charges within a tightly bound cell i, and $\Phi_2(i, j)$ is the interaction energy between charges in different tightly bound cells i and j. Because a ion-binding mode specifies only the number (instead of the specific locations) of the ions, we average over all the possible locations of the ions to compute the potential of mean force. In the calculations, to account for the solvent polarization effect, we apply the generalized Born (GB) model in the computation of $\Phi_1(i)$ and $\Phi_2(i, j)$ [8]-[10]; see Ref. [4, 6] for details.

ΔG_d in Eq. 2 contains three parts: the free energy for the electrostatic interactions between the diffusive ions and between the diffusive ions and the charges in the tightly bound region and the entropic free energy of the diffusive ions. ΔG_d is calculated by [1]-[6, 7]:

$$\Delta G_d = \frac{1}{2}\int \sum_\alpha c_\alpha(\mathbf{r}) z_\alpha q \left[\psi(\mathbf{r}) + \psi'(\mathbf{r})\right] d^3\mathbf{r} + k_\mathrm{B}T \int \sum_\alpha \left[c_\alpha(\mathbf{r})\ln\frac{c_\alpha(\mathbf{r})}{c_\alpha^0} - c_\alpha(\mathbf{r}) + c_\alpha^0\right] d^3\mathbf{r}, \tag{4}$$

where $\psi'(\mathbf{r})$ and $\psi(\mathbf{r})$ are the electrostatic potentials for system without and with the diffusive salt ions, respectively. $\psi'(\mathbf{r})$ is used here since $\psi(\mathbf{r}) - \psi'(\mathbf{r})$ gives the contribution to the potential from the diffusive ions. $\psi(\mathbf{r})$ and $\psi'(\mathbf{r})$ are obtained through solving the PB and the Poisson equations, respectively.

ΔG_b^{pol} in Eq. 2 is the (Born) self-polarization energy for the discrete charges in the tightly bound region [4], and is computed as the sum of Born (self-)energies of all the discrete charges (including the N phosphate charges and N_b tightly bound ions) inside the tightly bound region [4, 6]:

$$\Delta G_b^{pol} = \sum_i \Phi_0(i), \tag{5}$$

where $\Phi_0(i)$ is Born (self-) energy for charges inside the i-th tightly bound cell; see Ref. [4] for the detailed calculations.

Based on the above formulas, the electrostatic free energy for a RNA in ion solutions is computed by

$$G_\mathrm{E} = -k_B T \ln\left(Z/Z^{(id)}\right) = -k_B T \ln \sum_M \left(Z_M/Z^{(id)}\right), \tag{6}$$

and other electrostatic properties can also be computed through the partition functions.

Computation procedure and atomistic parameter sets

In the all-atomic TBI model, for an atomistic nucleic acid structure (e.g., from PDB or NDB data), the different types of atoms are treated as hard-core spheres with the radii determined by the respective van der Waals radii which has been tabulated in Table S1 in Supporting Material. Each phosphate (P) atom is considered to carry one unit negative charge at its center, while other types of atoms (C, O, N, etc) are treated as neutral spheres. The molecular van der Waals boundary defines a dielectric boundary: inside is continuous molecular interior with a low dielectric constant and outside is the continuous solvent with a high dielectric constant.

In the first step, for the given all-atom structure, we solve the PB equation to obtain the counterion distributions, from which we determine the tightly bound region for the multivalent counterions [1]-[4].

Second, we compute the pair-wise potentials of mean force $\Phi_1(i)$ and $\Phi_2(i, j)$ and Born energy $\Phi_0(i)$. In order to account for the dielectric discontinuity, we apply the generalized Born model for the calculations of the pair-wise charge-charge interactions and the self polarization energy of discrete charges [8]-[10]. The Born radii for discrete charges are estimated from the pair-wise method proposed by Hawkins [9], and the scaling parameters for atoms are taken the values in Ref. [10] which are tabulated according to the types of atoms; see Table S2. In the calculations for $\Phi_1(i)$, $\Phi_2(i, j)$ and $\Phi_0(i)$, the volume exclusion between ions and nucleic acid atoms are accounted for by a truncated Lenard-Jones potential:

$$\begin{aligned} U &= u_0\left(\frac{1}{r^{12}} - \frac{1}{r^6}\right) \quad \text{for } r < 1 \\ &\; 0 \qquad\qquad\qquad\quad\;\; \text{for } r > 1 \end{aligned}$$

where r is the distance between an ion and an atom in the unit of the sum of their radii. The parameter u_0 is taken as 0.35 since the exclusion between the hydrated ions and the molecular atoms are soft due to the soft H-atom exclusion. The calculated potentials of mean force $\Phi_1(i)$ and $\Phi_2(i, j)$ and the Born energy $\Phi_0(i)$ are tabulated and stored for the the calculations of the partition function in the next step.

Third, we enumerate the possible binding modes. For each mode, we calculate ΔG_b, ΔG_d, and ΔG_b^{pol} (Eq. 2). The sum over the binding modes gives the total partition function Z (Eq. 1) [1, 2, 6], from which we can calculate the electrostatic free energy.

We assume that ions are hydrated [1]-[6] and the ionic radii are equal to 3.5 Å for Na^+, 4.5 Å for Mg^{2+}, and 4 Å for K^+ [1]-[4, 11], respectively. We will also use a smaller divalent ion (with radius $\sim$ 3.5Å) for the purpose of examining the ion size effect in the PB calculations. Here, the dielectric constant ϵ of nucleic acid interior is set to be 20 [1, 4], and ϵ of solvent is set as the value of water [12]: $\epsilon(t) = 87.740 - 0.4008 \times t + 9.398 \times 10^{-4} \times t^2 - 1.41 \times 10^{-6} \times t^3$, where t is the temperature in Celsius. At 25°C, $\epsilon \simeq 78$. In the PB calculation, a thin charge-free layer of thickness of a cation radius is added to the molecular surface to account for the excluded volume layer of the cations [1]-[4]. The PB equation and the detailed ion distribution near the molecules are solved using the three-step focusing process [1]-[4]. For each run, the electrostatic potentials are iterated to a convergence of $< 10^{-4} k_B T/q$. The resolution of the first run varies with the grid size in order to make the iterative process viable, and the resolutions for the second and third runs are 1.36 Å per grid and 0.68 Å per grid, respectively. Our results are robust as tested against the different grid sizes.

Calculating $\Phi_1(i)$, $\Phi_2(i, j)$ and $\Phi_0(i)$ with the generalized Born model

As described in main text, we apply the generalized Born (GB) model to account for the RNA/solvent dielectric effects for the tightly bound ions. In this appendix, we show how to compute the pair-wise potential of mean forces $\Phi_1(i)$, $\Phi_2(i, j)$ and the polarization energy ΔG^{pol} with the generalized Born (GB) model [8]-[10]; see Eqs. 3 and 5 in Methods.

(1) $\Phi_1(i)$ and $\Phi_2(i, j)$ are calculated as the average over all the possible positions $\mathbf{R}$ of the tightly bound

ions in the respective tightly bound cells [1]-[6]:

$$\Phi_1(i) = -k_BT\ \ln\langle e^{-u_{ii}(\mathbf{R}_i)/k_BT}\rangle; \quad \Phi_2(i,j) = -k_BT\ \ln\langle e^{-u_{ij}(\mathbf{R}_i,\mathbf{R}_j)/k_BT}\rangle, \tag{7}$$

where u_{ii} is the electrostatic interactions for the charges in cell i and u_{ij} is the electrostatic interactions between the charges in two different cells i and j. In the calculations for $\Phi_1(i)$ and $\Phi_2(i,j)$, as shown below, we use the GB model to account for the polarization energy (due to dielectric discontinuity at the RNA/solvent interface) [8]-[10].

In the GB model, the electrostatic pair-wise interaction u_{ij} between two ($i \neq j$) charges q_i and q_j is given by

$$\begin{aligned} u_{ij} &= u_{ij}^{pol} + u_{ij}^{0}; && (8)\\ u_{ij}^{pol} &= -\left(\frac{1}{\epsilon_p} - \frac{1}{\epsilon_w}\right)\frac{q_iq_j}{\sqrt{d_{ij}^2 + \alpha_i\alpha_j\exp(-d_{ij}^2/(4\alpha_i\alpha_j))}}; && (9)\\ u_{ij}^{0} &= \frac{1}{\epsilon_p}\frac{q_iq_j}{d_{ij}}, && (10)\end{aligned}$$

where ϵ_p (= 20) and ϵ_w (= 78) are the dielectric constants of RNA interior and solvent, respectively, u_{ij}^{pol} is the polarization energy and u_{ij}^{0} is the Coulombic interaction energy in the uniform medium of dielectric constant ϵ_p. d_{ij} is the distance between the two charges. α_i and α_j are the Born radii for the two charges q_i and q_j.

(2) $\Phi_0(i)$ in Eq. 5 is the Born energy for charges inside the i-th tightly bound cell. For illustrative purpose, we assume there is one tightly bound ion in the i-th cell. $\Phi_0(i)$ is calculated from an averaging over all the possible positions $\mathbf{R}_i$ of the ion [4]:

$$\Phi_0(i) = -k_BT\ln\langle \mathrm{e}^{-(\Delta U_P^{pol}+\Delta U_I^{pol})/k_BT}\rangle, \tag{11}$$

where $\Delta U_P^{pol} = \Delta U_P^{pol}(i,\mathbf{R}_i)$ and $\Delta U_I^{pol} = \Delta U_I^{pol}(i,\mathbf{R}_i)$ are the self-energies of the phosphate i and of the ion (at position $\mathbf{R}_i$), respectively. $\langle ...\rangle$ denotes the averaging over all the possible ion positions $\mathbf{R}_i$ within the cell. With the GB approximation, we compute $\Delta U_P^{pol}(i,\mathbf{R}_i)$ and $\Delta U_I^{pol}(i,\mathbf{R}_i)$ using the following formulas [4]:

$$\Delta U_P^{pol}(i,\mathbf{R}_i) = -\left(\frac{1}{\epsilon_p} - \frac{1}{\epsilon_w}\right)\frac{q_P^2}{2\alpha_P(i,\mathbf{R}_i)}; \quad \Delta U_I^{pol}(i,\mathbf{R}_i) = -\left(\frac{1}{\epsilon_p} - \frac{1}{\epsilon_w}\right)\left(\frac{1}{\alpha_I(i,\mathbf{R}_i)} - \frac{1}{\alpha_I^0}\right)\frac{q_I^2}{2}. \tag{12}$$

where $\alpha_P(i,\mathbf{R}_i)$ and $\alpha_I(i,\mathbf{R}_i)$ are the Born radii for the phosphate i and the ion at $\mathbf{R}_i$, respectively. α_I^0 is the Born radius for an isolated ion.

(3) In the TBI model, we use a pair-wise method to calculate the Born radius α_i [9]-[10] for a charge i in the tightly bound region [4]:

$$\begin{aligned} \frac{1}{\alpha_i} &= \frac{1}{a_i} - \frac{1}{2}\sum_j A_j; && (13)\\ A_j &= \left(\frac{1}{L_{ij}} - \frac{1}{U_{ij}}\right) + \left(\frac{S_j^2a_j^2}{4d_{ij}} - \frac{d_{ij}}{4}\right)\left(\frac{1}{L_{ij}^2} - \frac{1}{U_{ij}^2}\right) + \frac{1}{2d_{ij}}\ln\frac{L_{ij}}{U_{ij}}, && (14)\end{aligned}$$

where

$$L_{\mathrm{ij}} = \begin{cases} 1 & \text{if } a_i \geq d_{ij} + S_j a_j; \\ \max(a_i, d_{ij} - S_j a_j) & \text{if } a_i < d_{ij} + S_j a_j, \end{cases} \tag{15}$$

and

$$U_{\mathrm{ij}} = \begin{cases} 1 & \text{if } a_i \geq d_{ij} + S_j a_j; \\ d_{ij} + S_j a_j & \text{if } a_i < d_{ij} + S_j a_j. \end{cases} \tag{16}$$

Here, $\sum_j$ denotes the sum over all the atoms, d_{ij} is the distance between charge i and atom j, a_i and a_j are the radii for the charge i and atom j. S_j is the structural scaling parameter and is equal to unity if there is no overlap between atoms. Generally, $S_j < 1$ for a realistic molecule. For different types of atoms in nucleic acids, the used scaling factor S_j are tabulated in Table S2 in the Supporting Material. For ions, the scaling parameter S is taken as 1 since there is no overlap between ions and nucleic acid molecules.

The Born radii for the charges (including phosphates and tightly bound ions) inside the tightly bound region are calculated with the method (Eq. 13) indicated here [4]: (1) For the Born radius of bare phosphates (without tightly bound ions), the summation in Eq. 13 is over all the atoms in RNA model. (2) For the Born radius of phosphates with tightly bound ions, the contribution of the bound ion is accounted for by including an additional term (A_j) in the summation (Eq. 13). (3) For the tightly bound ions, we calculate the Born radii on the grids. The Born radius for an ion at an arbitrary position in the tightly bound region can be approximated to be that on the closest grid point.

Table S1: van der Waals radii used in atomic TBI model[a]

atom types	P	C	N	O	H
van der Waals radii (Å)	1.9	1.7	1.6	1.5	1.0

[a] The van der Waals radii (Å) are taken from http://www.rbvi.ucsf.edu/chimera/current/docs/UsersGuide/midas/vdwtables.html.

Table S2: Scaling parameters used in atomic TBI model[a]

atom types	P	C	N	O	H
scaling parameters	0.86	0.72	0.79	0.85	0.85

[a] The scaling parameters for different types of atoms are taken from Ref. [10] in Supporting Material.

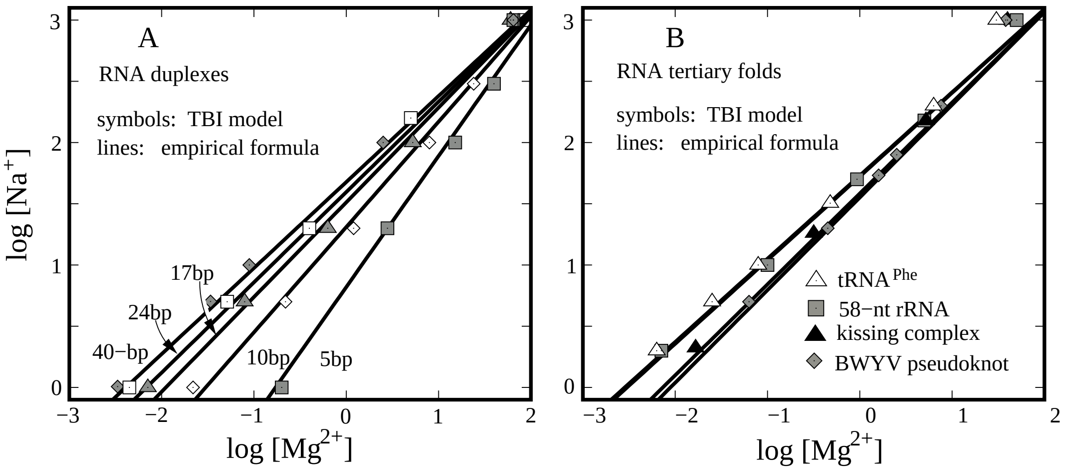

FIGURE S5 The equivalent relations between [Mg^{2+}] (in mM) and [Na^{+}] (in mM) in ionic neutralization for RNA molecules, i.e, $2f_{Mg^{2+}} \simeq f_{Na^{+}}$. (A) RNA duplexes with different lengths; (B) Four RNA tertiary folds: BWYV pseudoknot, 58-nt rRNA fragment, HIV-1 DIS kissing complex, and yeast $tRNA^{Phe}$. The lines are from the empirical formula (Eq. 4 in main text), and the symbols are the data from the TBI's calculations.

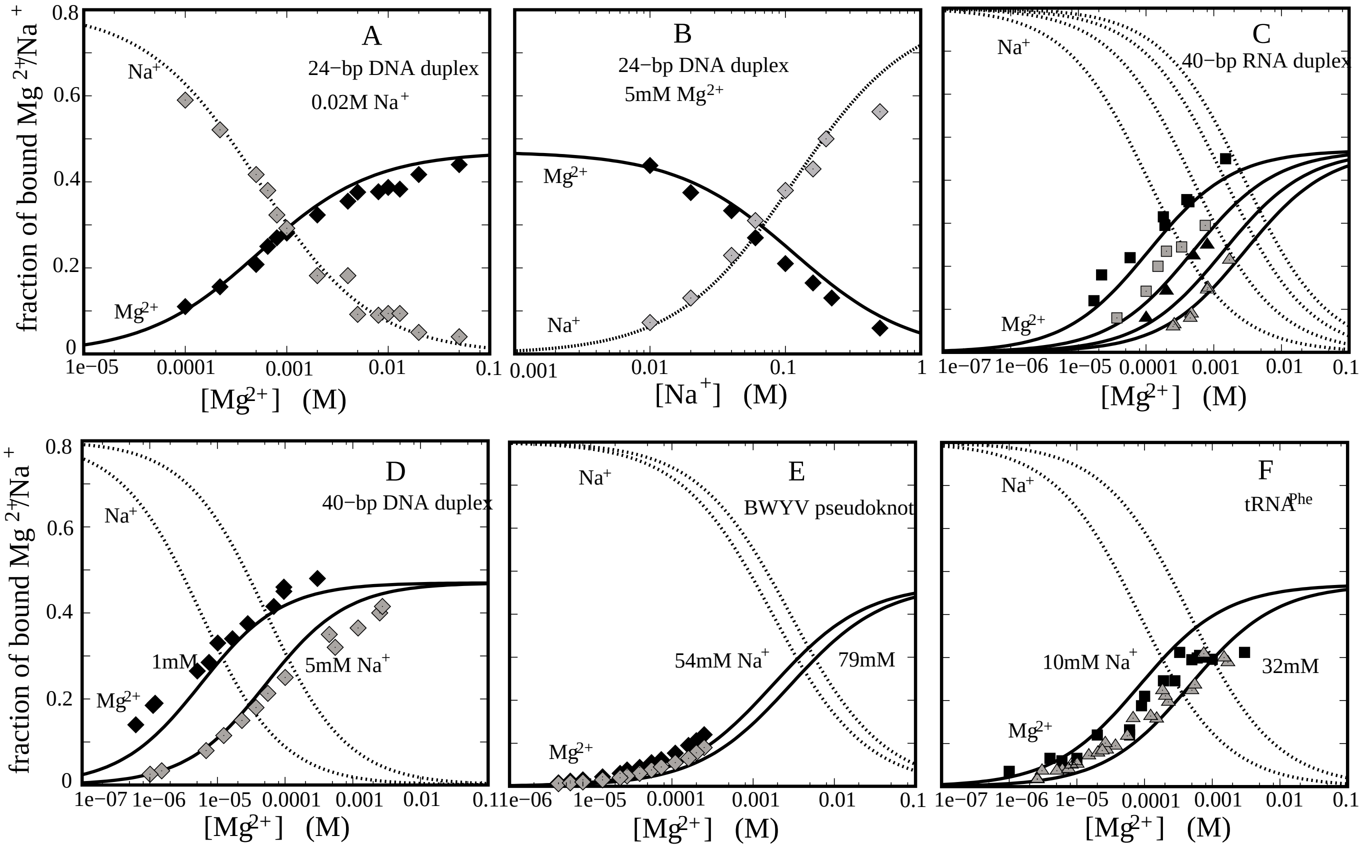

FIGURE S6 The Mg^{2+} and Na^{+} binding fractions per nucleotide for various nucleic acid molecules. The lines are from the empirical formulas (Eq. 6); and the symbols are experimental data. (A) 24-bp DNA duplex in 20mM Na^{+}. The experimental data are given in Fig. 2A; (B) 24-bp DNA duplex in 5mM Mg^{2+}. The experimental data are given in Fig. 2B; (C) 40-bp RNA duplex. The experimental data are for poly(A.U) and the reference is given in Figs. 2CD. From the left to right, $[Na^{+}]$=10mM, 29mM, 60mM, and 100mM, respectively; (D) 40-bp DNA duplex. The experimental data are for calf thymus DNA and are given in Figs. 2EF; (E) BWYV pseudoknot RNA. The experimental data are given in Figs. 3AB; (F) Yeast $tRNA^{Phe}$. The experimental data are given in Figs. 3EF.

References

[1] Tan, Z.J., and S.J. Chen. 2005. Electrostatic correlations and fluctuations for ion binding to a finite length polyelectrolyte. *J. Chem. Phys.* 122:044903.

[2] Tan, Z.J., and S.J. Chen. 2006. Ion-mediated nucleic acid helix-helix interactions. *Biophys. J.* 91:518-536.

[3] Tan, Z.J., and S.J. Chen. 2006. Electrostatic free energy landscape for nucleic acid helix assembly. *Nucleic Acids Res.* 34:6629-6639.

[4] Tan, Z.J., and S.J. Chen. 2008. Electrostatic free energy landscapes for DNA helix bending. *Biophys. J.* 94:3137-3149

[5] Chen, G., Z.J. Tan, and S.J. Chen. 2010. Salt dependent folding energy landscape of RNA three-way junction. *Biophys. J.* 98:111-120.

[6] Tan, Z.J. and S.J. Chen. 2009. Predicting electrostatic forces in RNA folding. *Methods Enzymol.* 469:465-487.

[7] Stigter, D. 1995. Evaluation of the counterion condensation theory of polyelectrolytes. *Biophys. J.* 69:380-388.

[8] Still, W.C., A. Tempczyk, R.C. Hawley, and T. Hendrickson. 1990. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* 112:6127-6129.

[9] Hawkins, G.D., C.J. Cramer, and D.G. Truhlar. 1995. Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.* 246:122-129.

[10] Liu, H.Y., I.D. Kuntz, and X. Zou. 2004. Pairwise GB/SA scoring function for structure-based drug design. *J. Phys. Chem. B* 108:5453-5462.

[11] Marcus, Y. 1985. Ion Solvation. John Wiley & Sons Ltd., Great Britain.

[12] Eisenberg, D., and W. Kauzmann. 1969. The Structure and Properties of Water. Oxford University Press, Oxford.